Prediction Framework for Statistical Respiratory Motion Modeling

Tobias Klinder¹, Cristian Lorenz², and Jörn Ostermann³

¹ Philips Research North America - Briarcliff Manor, NY, USA
tobias.klinder@philips.com

² Philips Research Europe - Hamburg, Germany

³ Institut für Informationsverarbeitung, Leibniz University of Hannover, Germany

Abstract. Breathing motion complicates many image-guided interventions working on the thorax or upper abdomen. However, prior knowledge provided by a statistical breathing model, can reduce the uncertainties of organ location. In this paper, a prediction framework for statistical motion modeling is presented and different representations of the dynamic data for motion model building of the lungs are investigated. Evaluation carried out on 4D-CT data sets of 10 patients showed that a displacement vector-based representation can reduce most of the respiratory motion with a prediction error of about 2 mm, when assuming the diaphragm motion to be known.

1 Introduction

Respiratory motion is a significant source of error for image guided interventions working on the thorax and upper abdomen. To reduce the uncertainties caused by breathing, different approaches for motion compensation have been presented, e.g., breathing coaching, patient-specific motion field extraction on the basis of pre-operatively acquired four-dimensional images, image-guided tracking throughout treatment or a combination of methods [1]. However, current approaches typically result in additional workload increasing treatment time.

One possibility to reduce breathing-induced problems without a significant increase in effort is by means of a statistical motion model, which is learned in advance. The prior knowledge about likely deformation can provide valuable information for image-guided interventions, image acquisition, but also to support image processing algorithms. Furthermore, the statistics covered in the model and thus the variability in breathing is interesting in itself.

Different motion models have been proposed in the past for cardiac motion [2,3] as well as for respiratory motion [4,5,6]. Besides the target organ, the motion models differ in the representation of the extracted data. However, different representations lead to different results and it is mostly not known in advance, which is best suited for the problem.

In this paper, a prediction framework for statistical lung motion modeling is proposed. Based on this general framework, different representations are presented and their characteristics and applicability for respiratory motion modeling

is investigated. The different models are compared by evaluating the model-based prediction of an unknown motion field.

Focus in this paper is set on lung motion modeling on the basis of 4D-CT. The created models are thereby patient-specific, i.e., the images for model building all belong to the same patient. As the purpose of this study was to analyse different model representations, patient-specific models are preferred over general motion models [4], due to the fact that the establishing of correspondence is much better defined. With a comparably high spatial resolution focus is set on 4D-CT.

2 Methods

2.1 Image Data

Free-breathing 4D-CT thorax data sets of 10 patients were acquired for radiotherapy treatment. In each treatment week, a ten-phase 4D-CT was acquired representing a complete breathing cycle. During acquisition of the 4D-CT data, patient respiratory traces were stored while the patient is breathing freely. The projection images were then retrospectively sorted into ten respiratory phase-based bins of 3D-CT image data based on the respiratory signal (i.e., from 0 % to 90 % phase at 10 % intervals with 0% being typically maximum inhale and 50 % maximum exhale). Depending on the course of treatment, 4D-CT data of up to eight weeks plus two weeks of preliminary examination was available per patient. For each patient at least nine 4D-CT scans were acquired. All images had an in-plane resolution of 0.85-0.97 mm and a slice thickness of 2.5 mm.

2.2 Motion Field Estimation

Lung motion fields are derived from the four-dimensional images using a surface-based tracking technique [7]. By propagating a topologically identical patient-specific lung surface mesh through all phases, anatomical point correspondences are assumed to be preserved. The trajectories of corresponding points of the adapted meshes thus provide a sparse motion field, which is finally interpolated using thin-plate-splines. The estimated continuous motion fields are denoted as $\mathcal{T}_{ij}$ defined between image i and j . We extensively evaluated the proposed method on lung 4D-CT images on the basis of publicly available data sets [8] achieving an accuracy of 1.0-1.7 mm in average.

For motion estimation, the lungs are at first segmented in one selected exhale image per patient. Afterwards patient-specific lung surface meshes are obtained by triangulating the thresholded image (cf. Fig. 1 (a)). The meshes cover the outer border of the lungs and also the surfaces of the pulmonary vessel tree as well as potential tumor surfaces [7]. By adapting the patient-specific reference lung mesh to the exhale images of all weeks of treatment, vertex correspondences are preserved for all images of the same patient. From the exhale phase, all meshes are tracked through the 4D-CT data set (cf. Fig. 1 (b) and (c)).

Fig. 1. Surface-based motion field estimation. Patient-specific lung mesh (a) tracked through all phases, exemplarily shown for end-exhale (b) and end-inhale (c). Note that inside structures are covered in the mesh.

2.3 Intra-week Alignment

The 4D-CT data are acquired at different treatment days. For statistical modeling, the data has to be aligned. To remove differences in patient position and image orientation but to remain differences in breathing level, all 4D-CT data sets are at first aligned by means of rigid transformations based on the segmented spinal column.

2.4 Statistical Motion Model

Statistical modeling is based on expressing the extracted data as a random vector $\mathbf{z}$ with different possible representations proposed in the next section. A set of N corresponding vectors $\mathbf{z}_n$ from a learning set, e.g., obtained from N different 4D-CT data sets, is combined in a matrix $\mathbf{Z} = [\mathbf{z}_1 | \mathbf{z}_2 | \dots | \mathbf{z}_N]$ and statistics can be applied on $\mathbf{Z}$. Note that the rows of the individual vectors $\mathbf{z}_n$ contain corresponding information.

From the learning set, the mean is given as $\bar{\mathbf{z}} = \frac{1}{N} \sum_{n=0}^{N-1} \mathbf{z}_n$ and higher order statistics are calculated using the covariance matrix of the data

$$\Sigma_{\mathbf{ZZ}} = \tilde{\mathbf{Z}} \tilde{\mathbf{Z}}^T, \quad (1)$$

which gives the first order *statistical motion model*. The matrix $\tilde{\mathbf{Z}}$ contains the learning set after subtracting the mean.

2.5 Data Representation

The extracted information extracted from the four-dimensional images covers: i) adapted surface meshes in each phase, where each mesh is given as a set of K vertices building a vector $\mathbf{s} = (x_1, y_1, z_1, x_2, y_2, z_2, \dots, x_K, y_K, z_K)^T$, ii) deformations $\mathcal{T}_{ij}$ between the images of phase i and j , iii) volume of the lungs V obtained from a lung segmentation.

From this extracted information different representations can be defined that are based on the absolute vertex locations of the segmentations or on the displacement vectors. Furthermore, time can be modeled implicitly based on the

temporal binning of the data (e.g., for 4D-CT the binning into ten phases), or explicitly. Four representations are given in the following.

Shape-Based. The shape-based representation takes the absolute vertex positions provided in all images of a dynamic sequence. Given a ten-phase 4D-CT, the vector $\mathbf{z}_{SM}$ can be defined as

$$\mathbf{z}_{SM} = [\mathbf{s}_0, \mathbf{s}_1, \dots, \mathbf{s}_8, \mathbf{s}_9]^T. \quad (2)$$

where $\mathbf{s}_i$ is the segmentation i -th phase.

Volume-Binned Shape-Based. In the case of the $\mathbf{z}_{SM}$, temporal corresponding phases are correlated in the statistical model that nevertheless can show high differences in shape due to differences in breathing. The idea of the volume-based representation is to correlate shapes that are more similar by replacing the temporal binning of the data by a volume binning (cf. Fig. 2 (a)). At first, the patient-specific volume range defined by a maximal and minimal volume is separated into a predefined number of bins. Then, the adapted meshes are sorted by there corresponding breathing volume as

$$\mathbf{z}_{VOL_i} = [\mathbf{s}], V_{li} \leq V < V_{ui}, \quad (3)$$

where V_{li} and V_{ui} define the lower and upper volume range of the i -th volume bin. The statistical model is then built for each bin from all samples over the weeks of treatment that belong to the corresponding volume range. As this representation cannot express the effect of hysteresis, separate statistical ensembles have to be built for inhalation and exhalation. Assuming K bins, this results in K models for inhalation and exhalation, respectively.

Displacement Vector-Based. Both proposed representations are so far based on absolute landmark locations. Alternatively, the representation can also be based on the displacement field. Motion field extraction results in a set of continuous transformation $\mathcal{T}_{ij}$ that explain the deformation between the images. In the case of a ten-phase 4D-CT, $\mathcal{T}_{ij}$ is given with respect to a reference phase (commonly the end-exhale phase which is denoted as phase 5). Thus, the $\mathcal{T}_{ij}$, become $\mathcal{T}_{5j}$ with $j = 0, 1, \dots, 4, 6, \dots, 9$. The displacement vector-based representation $\mathbf{z}_{MF}$ contains the deformations defined at a set of K vertices. In contrast to the continuous transformation $\mathcal{T}$, a motion field defined at a given set of locations is denoted as $\mathbf{m} = [\delta x_1, \delta y_1, \delta z_1, \dots, \delta x_K, \delta y_K, \delta z_K]^T$ where δx_k , δy_k , δz_k give the displacement in $\mathbf{x}$ -, $\mathbf{y}$ -, and $\mathbf{z}$ -direction of the k -th vertex location. The vector $\mathbf{z}_{MF}$ is then the concatenation, e.g., if ten phases should be taken into account

$$\mathbf{z}_{MF} = [\mathbf{m}_{50}, \mathbf{m}_{51}, \dots, \mathbf{m}_{54}, \mathbf{m}_{56}, \dots, \mathbf{m}_{59}]^T. \quad (4)$$

The proposed vector is based on differences and can be considered as a deformation model (cf. Fig. 2 (b)). It represents the deformation of a discrete set of landmarks. Thus, for motion prediction, reference locations have to be provided. Compared to $\mathbf{z}_{SM}$, the vector $\mathbf{z}_{MF}$ contains $3 \times K$ less entries.

Parametric Displacement Vector-Based. The vector $\mathbf{z}_{MF}$ models the time implicitly. The temporal binning of the model is defined by the temporal binning of the data. Another way to represent the trajectory is by means of a parametrisation of the trajectory. By describing the displacement over time along one coordinate as a discrete signal r_k , the discrete fourier transform $R_u = \frac{1}{M} \sum_{k=0}^{M-1} r_k e^{-j2\pi uk/M}$ can be calculated, where R_u are called fourier coefficients. The trajectory can be described with less parameters when using only the first $L \leq M$ complex coefficients. As the input signal is real, the first fourier coefficient is real. Thus, taking exemplarily only the first two fourier coefficients, the vector $\mathbf{z}_{DFT}$ is defined as

$$\mathbf{z}_{DFT} = [Re\{R_{0,x1}\}, Re\{R_{1,x1}\}, Im\{R_{1,x1}\} \dots, Re\{R_{0,zK}\}, Re\{R_{1,zK}\}, Im\{R_{1,zK}\}]^T, \quad (5)$$

where $Re\{R_{1,xk}\}$ and $Im\{R_{1,xk}\}$ are the real and imaginary part of the second complex fourier coefficient in $\mathbf{x}$ -direction at the k -th vertex position. Although loosing some accuracy compared to the true trajectory, this representation describes the trajectory continuously and reduces the number of parameters per location (cf. Fig. 2 (c)). Due to the fact that the trajectory over the breathing cycle is rather elliptical shaped, it is mostly sufficient to choose $L = 2$ or $L = 3$.

Fig. 2. Different model representations. The data is sorted in (a) into bins based on the breathing volume which is done separately for inhalation and exhalation. In (b), motion fields from a reference phase are calculated to the individual phases. The trajectory over the breathing cycle is described in (c) by fourier coefficients $R_0, \dots, R(M)$ as a continuous curve.

2.6 Prediction

Independant of the exact representation, the basic idea of the prediction framework is to use the statistics covered in $\mathbf{Z}$ to predict a patient-specific motion, e.g., at a new treatment day. For that purpose, the vector $\mathbf{z}$ will be partitioned into two parts $\mathbf{z} = [\frac{\mathbf{y}}{\mathbf{x}}]$, where $\mathbf{y}$ gives the part to be predicted and the $\mathbf{x}$ give the

information available during application to drive the model. Motion prediction is formulated as a linear multivariate regression problem

$$\mathbf{y}_{\text{pred}} = \mathbf{B} \cdot \mathbf{x}, \quad (6)$$

where $\mathbf{B}$ can be found from the learning set using ordinary least squares

$$\text{argmin}(\mathbf{Y} - \mathbf{B}\mathbf{X})^2 \Rightarrow \mathbf{B} = \mathbf{Y}\mathbf{X}^T(\mathbf{X}\mathbf{X}^T)^{-1} = \boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1} \quad (7)$$

where $\boldsymbol{\Sigma}_{XX}$ is the covariance matrix of $\mathbf{X}$ and $\boldsymbol{\Sigma}_{YX}$ the cross-covariance matrix of $\mathbf{Y}$ and $\mathbf{X}$. However, while the solution is optimal in a least squares sense, it is highly sensitive to overfitting. When the dimension of $\mathbf{X}$ is higher than the number of observations, $\mathbf{X}\mathbf{X}^T$ is likely to be singular and the regression approach is no longer feasible, i.e., because of multicollinearity. We therefore perform Principal Component Regression (PCR) to reduce the noise contained in the matrix $\boldsymbol{\Sigma}_{XX} = \mathbf{X}\mathbf{X}^T$ by expressing $\boldsymbol{\Sigma}_{XX}$ as

$$\boldsymbol{\Sigma}_{XX} = \sum \phi_p \lambda_p \phi_p^T, \quad (8)$$

where ϕ_p is the p -th eigenvector with eigenvalue λ_p . By taking only the first P principal components that belong to the P largest eigenvalues, only the directions of largest variance are considered preventing the singularity of $\boldsymbol{\Sigma}_{XX}^{-1}$ [9].

3 Results

The different representations are evaluated by using the respective models for motion prediction. As a motion stimulator that drives the model, motion information at a small part close to the diaphragm was assumed to be known, as shown in Fig. 3. From our point of view, using a statistical motion model for motion prediction to support image guided interventions is one of the most relevant scenarios. The diaphragm is chosen as it is the main breathing motor and it can be easily tracked in many imaging modalities.

Evaluation is carried out in a leave-one out manner, i.e., the model is built from all weeks but the one under consideration. For the shape-based representations, the prediction result is given in the form of absolute landmark locations and the L_2 -norm between predicted and propagated vertices is calculated. The displacement vector-based prediction is evaluated by calculating the L_2 -norm between predicted and estimated motion vectors. For PCR regression, the first $N/2$ of N possible principal components were chosen which resulted in best results.

The results for the different representations are shown in Fig. 4 (a) and (b). As for the displacement vector-based prediction, the end-exhale was chosen as

Fig. 3. Labelled lung meshes. Motion information at small parts of the diaphragm (blue) assumed to be known.

Fig. 4. Results of motion prediction in end-inhale phase over weeks of treatment for (a) shape-based representation and (b) displacement-vector based representation. In addition, mean lung amplitude (LM) is given.

the reference phase. Evaluation was focused on the error in the end-inhale phase being the phase with largest distance. In both Fig. 4 (a) and (b), the mean motion amplitude from end-exhale to end-inhale is given for comparison.

The results of the two shape-based representations, $\mathbf{z}_{SM}$ and $\mathbf{z}_{VOL_i}$ with $K = 5$ are given in Fig. 4 (a). Both representations can compensate most of the motion amplitude. However, setting the number K of volume bins, turned out to be complicated. If the bin size is too small, only few samples might be available for a certain volume and if it is too large, shapes with very different volumes will be correlated. As $\mathbf{z}_{VOL_i}$ is based on the volume, it can nevertheless be easily adapted to data with a different temporal binning. The results of the two displacement vector-based representations, $\mathbf{z}_{MF}$ and $\mathbf{z}_{DFT}$ using three complex fourier coefficients, is shown in Fig. 4 (b). As both representations are based on the displacement from exhalation to inhalation, reference locations of the corresponding exhalation state have to be provided. Due to this additional information, the error is much lower compared to a shape-based representation. Although $\mathbf{z}_{DFT}$ provides a more compact representation of the trajectory, the results are slightly worse compared to $\mathbf{z}_{MF}$. This was caused by the fact that the approximation using three fourier coefficients introduced an additional error of 0.40 mm in average. However, the advantage of $\mathbf{z}_{DFT}$ lies in the continuous representation. Both $\mathbf{z}_{MF}$ and $\mathbf{z}_{DFT}$ achieved an error of about 2 mm.

4 Conclusion

Prior knowledge provided by a statistical motion model can help to overcome the problems caused by breathing. The main contributions of this paper are i) the design of a general framework for motion prediction and ii) the proposition and comparison of different representations. In addition to representations that are based on the temporal binning of the data, more flexible representations are

proposed that describe the trajectory continuously or that are oriented on the breathing volume. In principle, adaptation to data, that does not show a fixed temporal binning is thus possible. For motion prediction, the diaphragm motion was assumed to be known. The respective models were evaluated by comparing predicted and estimated motion fields. Although breathing motion is known to be highly variable, it was in each case possible to compensate for most of the breathing amplitude. Different representations showed different characteristics and it thus depends on the clinical application and the available data which representation fits best. In general, deformation models that are based on the displacement vectors perform better than those that are based on the absolute landmark locations. On the basis of 4D-CT data sets of 10 patients, a prediction error of about 2 mm could be obtained when using a displacement vector-based representation. Although the framework was evaluated on the basis of 4D-CT, it can be easily adapted to other modalities as, e.g., 4D-MR.

References

1. Keall, P., Mageras, G., Balter, J., et al.: The management of resp. motion in radiation oncology report of AAPM Task group. *Med. Phys.* 33(10), 3874–3900 (2006)
2. Bosch, J., Mitchell, S., Lelieveldt, B., et al.: Automatic segmentation of echocardiographic sequences by active appearance motion models. *IEEE Transactions on Medical Imaging* 21 (11), 1374–1383 (2002)
3. Casero, R., Noble, J.: A novel explicit 2D+t cyclic shape model applied to echocardiography. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) *MICCAI 2008, Part I. LNCS*, vol. 5241, pp. 527–534. Springer, Heidelberg (2008)
4. Ehrhardt, J., Werner, R., Schmidt-Richberg, A., et al.: Generation of a mean motion model of the lung using 4D-CT image data. In: *Eurographics Workshop on Visual Computing for Biomedicine* (2008)
5. McClelland, J., Blackall, J., Tarte, S., et al.: A continuous 4D motion model from multiple respiratory cycles for use in lung radiotherapy. *Med. Phys.* 33 (9), 3348–3358 (2006)
6. Zhang, Q., Pevsner, A., Hertanto, A., et al.: A patient-specific respiratory model of anatomical motion for radiation treatment planning. *Medical Physics* 34(12), 4772–4782 (2007)
7. von Berg, J., Barschdorf, H., Blaffert, T., et al.: Surface based cardiac and respiratory motion extraction motion extraction for pulmonary structures from multi-phase CT. In: *Proc. of SPIE*, vol. 6511, pp. 65110Y1–11 (2007)
8. Castillo, R., Castillo, E., Guerra, R., et al.: A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets. *Physics in Medicine and Biology* 54, 1849–1870 (2009)
9. Rosipal, R., Krämer, N.: Overview and recent advances in partial least squares. In: Saunders, C., Grobelnik, M., Gunn, S., Shawe-Taylor, J. (eds.) *SLSFS 2005. LNCS*, vol. 3940, pp. 34–51. Springer, Heidelberg (2006)